

El BSC hace pública la version 2.9 de PyCOMPSs/COMPSs y la versión 0.7.0 de la dislib



El Barcelona Supercomputing Center ofrece a la comunidad HPC un conjunto de herramientas que ayudan a los desarrolladores a programar y ejecutar sus aplicaciones de forma eficiente en infraestructuras computacionales distribuidas. En particular, PyCOMPSs ofrece la interface para aplicaciones Python que es utilizada como herramienta para la convergencia de la supercomputación, inteligencia artificial y big data.

Esta versión de COMPSs incluye soporte para servicios REST, tareas "request" a servicios http que permitirán un nuevo tipo de flujos de trabajo que combinan tareas tradicionales y tareas http. El mecanismo también puede utilizarse con tareas de Función como servicio (FaaS) predesplegadas.

Para las aplicaciones en Python, PyCOMPSs incluye una nueva función de análisis de los accesos a los datos que permite decidir qué parámetros tienen una tasa de reutilización más alta y obtener beneficios de almacenarlos en la caché.

La nueva versión de la dislib incluye una versión de la descomposición QR a bloques con tres modos, regresores de random forest paralelos, y una multiplicación de matrices que permite realizar la transpuesta de los argumentos in-situ.

El grupo [Workflows and Distributed Computing](#) del Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC) está orgulloso de anunciar el lanzamiento de la versión 2.9 (nombre en clave Jasmine), del entorno de programación [COMPSs](#).

Esta versión de COMPSs, disponible a partir de hoy, actualiza el resultado del trabajo del equipo en los últimos años en la provisión de un conjunto de herramientas que ayudan a los desarrolladores a programar y ejecutar sus aplicaciones de forma eficiente en infraestructuras computacionales distribuidas, como clústers, nubes y clústeres gestionados con contenedores. COMPSs es un modelo de programación basado en tareas capaz de mejorar notablemente el rendimiento de aplicaciones a gran escala mediante la paralelización de manera automática de su ejecución.

COMPSs ha estado disponible en los últimos años por los usuarios del supercomputador MareNostrum y en la Red Española de Supercomputación (RES) y ha sido usado en varios proyectos de investigación como EUBra-BIGSEA, MUG, EGI, ASCETIC, TANGO, NEXTGenIO, I-BiDaaS, mF2C y CLASS. En estos proyectos se ha utilizado COMPSs para desarrollar casos de uso proporcionados por diferentes comunidades de diversas disciplinas como la biomedicina, la ingeniería, la biodiversidad, la química, la astrofísica, finanzas, telecomunicaciones, procesos de fabricación y ciencias de la tierra. Actualmente también está en extensión y en uso en aplicaciones en los proyectos AI-SPRINT, ExaQUte, LANDSUPPORT, el BioExcel CoE, PerMedCoE, ELASTIC, i en el proyecto Edge Twins HPC del tipo Innovation Launchpad. También ha sido aplicado a un caso de uso del CHEESE CoE. Una mención especial es el proyecto eFlows4HPC coordinado por el grupo, iniciado en enero de 2021, que tiene como objetivo desarrollar una pila de software para workflows donde uno de los componentes principales es el entorno PyCOMPSs / COMPSs.

La nueva versión incluye soporte para servicios REST en forma de tareas "request" a servicios http, tanto para aplicaciones Python como Java. El mecanismo también puede utilizarse con tareas de Función como servicio (FaaS) predespliegadas. Este soporte mejora el soporte anterior de COMPSs a los servicios web que se limitaba a los servicios web SOAP y a las aplicaciones Java COMPSs. La sintaxis se ha ampliado con anotaciones para las tareas http, así como con el soporte de ejecución COMPSs correspondiente. Esta funcionalidad permite un nuevo tipo de flujos de trabajo que puede combinar tareas tradicionales con el código proporcionado por el desarrollador de la aplicación, con invocaciones a tareas http externas.

La versión anterior 2.9 proporcionaba soporte para una memoria compartida para las tareas Python para evitar el tiempo de serialización y deserialización de los argumentos. La versión actual amplía esta funcionalidad con un mecanismo de apoyo a la decisión para los desarrolladores que hace un análisis de qué parámetros de las tareas deben almacenarse en la caché y cuáles no. Dado que el almacenamiento en la memoria caché comporta una sobrecarga adicional, el mecanismo proporciona información sobre la reutilización de los parámetros en términos de acceso a la memoria caché que permite al desarrollador inferir cuáles se beneficiarán de ser almacenados en la memoria caché.

Otras mejoras son la extensión del soporte para tareas MPI con el parámetro `processes_per_node` que permite la distribución de procesos a los nodos, soporte para versiones de Java superiores a 8, soporte parcial para sistemas macOS y soporte de serialización de objetos PyArrow. Además, la versión 2.10 de COMPSs viene dotada con otras funcionalidades menores, extensiones y resolución de errores.

COMPSs ha tenido alrededor de **1000 descargas** durante el año pasado y es utilizado por unos **20 grupos** en aplicaciones reales. Recientemente, COMPSs ha atraído el interés de áreas como la ingeniería, el reconocimiento de imágenes, la genómica y sismología, áreas en las que se han realizado cursos específicos y acciones de divulgación.

Los paquetes y la lista completa de funciones están disponibles en la página de [descargas](#). En esta página también se puede encontrar una máquina virtual que permite probar las funcionalidades de COMPSs a través de un tutorial paso a paso que guía al usuario en el desarrollo y ejecución de un conjunto de aplicaciones de ejemplo.

Además del software, hay un conjunto de guías de usuario y administrador, y artículos publicados en conferencias y revistas relevantes.

Para más información, visite nuestra página web: <http://www.bsc.es/compss>



El grupo también anuncia la nueva versión de dislib 0.7.0. La biblioteca de computación distribuida (dislib) proporciona algoritmos distribuidos preparados para utilizarlos como biblioteca. Hasta ahora, dislib se ha centrado en algoritmos de aprendizaje automático y ofrece una interfaz inspirada en scikit-learn. El objetivo principal de dislib es facilitar la ejecución de algoritmos de análisis de big data en plataformas distribuidas, como clusters, nubes y superordenadores. Dislib se ha implementado utilizando el modelo de programación PyCOMPSs.

Dislib se basa en una estructura de datos distribuidos, ds-array, que permite la ejecución paralela y distribuida de los métodos de aprendizaje automático. El código de la biblioteca dislib se implementa como una aplicación PyCOMPSs, donde los distintos métodos se anotan como tareas PyCOMPSs. En el momento de la ejecución, PyCOMPSs se encarga de todos los aspectos de paralelización y distribución de datos. Sin embargo, el código de usuario del usuario final de la dislib omite los aspectos de paralelización y distribución, y está escrito como simples scripts de Python, con una interfaz muy parecida a la interfaz de scikit-learn. Dislib incluye métodos de agrupación, clasificación, regresión, descomposición, selección de modelos y gestión de datos. Un contrato de investigación con FUJITSU había financiado parcialmente la biblioteca dislib y se utilizó para evaluar al procesador A64FX. Actualmente, los desarrollos dislib están cofinanciados en un 50% por el Fondo Europeo de Desarrollo Regional de la Unión Europea en el marco del Programa Operativo FEDER de Cataluña 2014-2020, por el proyecto H2020 AI-Sprint y por el proyecto EuroHPC eFlows4HPC .

Desde su creación, dislib se ha aplicado en casos de uso de astrofísica (DBSCAN, con datos de la misión GAIA), análisis de resultados de dinámicas moleculares (Daura y PCA, BioExcel CoE). En el proyecto eFlows4HPC, se está aplicando en dos casos de uso: en informática urgente para riesgos naturales (regresores random forest) y en gemelos digitales para la fabricación (QR).

La nueva versión 0.7.0 incluye una implementación paralela de los regresores random forest y la descomposición QR en bloques con tres modos: completo, económico y r. La implementación de QR se basa en reflectores Householder con el enfoque inspirado en las rotaciones de Givens proporcionando un mejor paralelismo. La nueva versión también contiene una implementación de MinMax Scaler para el preprocesamiento de los datos de entrada, nuevas funciones de utilidad para matrices para eliminar las últimas filas o columnas, y una multiplicación de matrices con transpuesta de las matrices in-situ. También se han realizado diversas mejoras de rendimiento en otros algoritmos. Además de esto, todas las tareas implementadas en dislib a partir de ahora pueden utilizar un número especificado de procesadores.

Dislib 0.7.0 viene con otras extensiones y con una nueva guía de usuario. El código es abierto y está disponible para su [descarga](#).

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 27 Dic 2024 - 00:50): <https://www.bsc.es/es/noticias/noticias-del-bsc/el-bsc-hace-p%C3%BAblica-la-version-29-de-pycompsscompss-y-la-versi%C3%B3n-070-de-la-dislib>